

QUALITY ASSESSMENT TOOL

This invention relates to a non-intrusive speech quality assessment system.

5

Signals carried over telecommunications links can undergo considerable transformations, such as digitisation, encryption and modulation. They can also be distorted due to the effects of lossy compression and transmission errors.

10

Objective processes for the purpose of measuring the quality of a signal are currently under development and are of application in equipment development, equipment testing, and evaluation of system performance.

15

Some automated systems require a known (reference) signal to be played through a distorting system (the communications network or other system under test) to derive a degraded signal, which is compared with an undistorted version of the reference signal. Such systems are known as "intrusive" quality assessment systems, because whilst the test is carried out the channel under test cannot, in general, carry live traffic.

20

Conversely, non-intrusive quality assessment systems are systems which can be used whilst live traffic is carried by the channel, without the need for test calls.

25

Non-intrusive testing is required because for some testing it is not possible to make test calls. This could be because the call termination points are geographically diverse or unknown. It could also be that the cost of capacity is particularly high on the route under test. Whereas, a non-intrusive monitoring application can run all the time on the live calls to give a meaningful measurement of performance.

10 A known non-intrusive quality assessment system uses a database of distorted samples which has been assessed by panels of human listeners to provide a Mean Opinion Score (MOS).

15 MOSs are generated by subjective tests which aim to find the average user's perception of a system's speech quality by asking a panel of listeners a directed question and providing a limited response choice. For example, to determine listening quality users are asked to rate "the
20 quality of the speech" on a five-point scale from Bad to Excellent. The MOS, is calculated for a particular condition by averaging the ratings of all listeners.

In order to train the quality assessment system each
25 sample is parameterised and a combination of the parameters is determined which provides the best prediction of the MOSs indicted by the human listeners. International Patent Application number WO 01/35393 describes one method for paramterising speech samples for
30 us in a non-intrusive quality assessment system.

This invention relates to improved parameters for assessing speech quality over a packet switched network, in particular over Voice Over Internet Protocol (VOIP) networks.

5

According to the invention there is provided a method and apparatus for storing a sequence of intercepted packets associated with a call each packet containing speech data, and an indication of a transmission time of said packet;
10 storing with each intercepted packet an indication of an intercept time of said packet; extracting a set of parameters from said sequence of packets; and generating an estimated mean opinion score in dependence upon said set of parameters; characterised in that the extracting
15 step comprises the sub steps of: generating a jitter parameter for each of a sequence of stored packets in dependence upon the difference between the transmission time of a stored packet and the transmission time of a preceding stored packet of the sequence; and the
20 difference between the intercept time of said stored packet and the intercept time of said preceding packet; generating a long term average jitter parameter for said stored packet in dependence upon the value of said jitter parameter for said stored packet and the value of said
25 jitter parameter for any preceding stored packets; and generating a differential jitter parameter in dependence upon the jitter parameter and the long term jitter differential parameter.

30

Embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings, in which:

5 Figure 1 is a schematic illustration of a non-intrusive quality assessment system;

 Figure 2 is a block diagram illustrating a non-intrusive quality assessment system monitoring calls
10 between an IP network and a circuit switched network;

 Figure 3 is a block diagram of a VOIP gateway;

 Figure 4 is a block diagram illustrating functional
15 block of an apparatus for quality assessment;

 Figure 4a is a flow chart illustrating the steps carried out by the apparatus of Figure 4;

20 Figure 5 is an illustration of parameters produced by a parameterisation process;

 Figure 5a is a flow chart showing abroad overview of a parameterisation process;
25

 Figure 6 illustrates combination of parameters at various levels;

 Figure 7 illustrates use of a sliding window; and
30

Figure 8 is a flow chart illustrating calculation of a particular parameter;

Referring to Figure 1, a non-intrusive quality assessment
5 system 1 is connected to a communications channel 2 via an
interface 3. The interface 3 provides any data conversion
required between the monitored data and the quality
assessment system 1. A data signal is analysed by the
quality assessment system, as will be described later and
10 the resulting quality prediction is stored in a database
4. Details relating to data signals which have been
analysed are also stored for later reference. Further data
signals are analysed and the quality prediction is updated
so that over a period of time the quality predication
15 relates to a plurality of analysed data signals.

The database 4 may store quality prediction results
resulting from a plurality of different intercept points.
The database 4 may be remotely interrogated by a user via
20 a user terminal 5, which provides analysis and
visualisation of quality prediction results stored in the
database 4.

Referring now to Figure 2, a VOIP gateway 40 converts data
25 at an interface between a circuit switched network 20 and
an IP network 26. The IP network 26 comprises a plurality
of IP routers 46. A VOIP probe 10 monitors VOIP calls to
assess quality of speech provided by the IP network.

30 VOIP can be divided into two broad system types; systems
that transport voice over the Internet and systems that

carry voice across a managed IP network.

The VOIP packet stream itself is well defined so VOIP calls can be identified either by monitoring call control
5 signalling and extracting call set-up messages or by being able to recognise VOIP packets. The probe 10 of the present invention recognises VOIP packets as this enables calls to be identified even if the start of the call is missed. This technique also avoids problems when the
10 packet stream and signalling information travel via different routes.

In order to monitor the speech quality of a VOIP from within the IP network, there is a need to account for the
15 highly non-linear VOIP gateway 40.

The probe 10 needs to account for each gateway according to the properties of the gateway because different gateway implementations respond to the effects of IP transmission
20 in varying ways.

Figure 3 illustrates a simple VIOP gateway 40. A jitter buffer 41 receives an IP packet stream. The jitter buffer 41 removes jitter and re-orders any mis-sequenced packets.
25 The packets are then sent to a speech decoder 42 in the appropriate time sequence where they are decoded.

An error concealer 43 uses error concealment techniques to mask any missing packets to provide an audio signal.

30

There are numerous VOIP gateway manufacturers - each produces a number of different gateways, each one operating slightly differently. It would be ideal if all of these gateways could be assumed to produce the same speech quality output from a given IP packet stream - but in fact different gateways will produce different speech quality scores from the same IP packet stream.

For example, a single manufacturer may use a variety of different jitter buffer algorithms for the jitter buffer 41. The impact on speech quality of the jitter buffer is heavily dependent on the effectiveness of a specific algorithm and implementation.

Speech decoders are generally standardised and well known. However, the effects of additional error concealment when encountering lost packets vary. Both jitter buffer and error concealment algorithms tend to be proprietary and can vary widely from gateway to gateway.

Therefore to accurately predict a speech quality MOS from an IP packet stream (or even a post jitter-buffer packet stream) non-intrusive predictors, such as the VOIP probe 10 of the present invention, need to take account of the specific gateway in use.

The probe 10 is calibrated for each different type of VOIP gateway which is supported. The calibration process involves characterising a gateway's speech quality performance over a wide range of network conditions. Once

a gateway has been characterised this information is stored in a calibration file, which can be loaded on command into the probe 10 and used to achieve highly accurate quality monitoring.

5

If a gateway is used which has not been calibrated then the probe 10 can still be used. However, in this case the output may not be representative of a MOS.

10 The probe 10 will now be described in more detail with reference to Figure 4 and Figure 4a. Figure 4 illustrates means for performing a quality assessment process, and Figure 4a illustrates the method steps to be carried out by the apparatus of Figure 4.

15

Capture module 50 at step 70 captures and stores an IP packet, and records the time of capture. Any corrupt packets are discarded. A call identification module 52 identifies to which call a captured packet belongs at
20 step 72. A pre-process module 54 discards any information from the captured packet which is no longer needed at step 74, in order to reduce memory and processing requirements for subsequent modules.

25 A resequence buffer 56 is used to store packet data, and to either pass the data to subsequent modules in sequence, or provide an indication that the data did not arrive at the correct time at step 76. The resequence buffer 56 used in this embodiment of the invention is a simple cyclic
30 buffer.

A voice activity detector 58 labels each packet as either speech or silence at step 78. 'Missing' packets are classified to the same classification as the immediately preceding packet.

Parameterisation module 60 extracts parameters from the packet data at step 80 in order to provide a set of parameters which are indicative of the likely MOS for the speech signal carried by the sequence of packet data associated with a particular call.

A prediction module 62 is then used to predict the MOS at step 82 based on a sequence of parameters received from the parameterisation module 60. A MOS will not be calculated until a predetermined number of packets associated with a particular monitored call have been received.

The parameterisation module will now be described with reference to Figures 5 to 8.

Parameters which are used for a particular gateway are defined within the calibration file. Parameters are calculated as follows. Every time new packet data is received from the VAD module 58 basic parameters are calculated. These basic parameters are combined over time in various ways to calculate 'level two' parameters. The level two parameters are then used to calculate 'level three' parameters.

Figure 5 and Figure 5a broadly illustrate this process. For example, when packet data (number 5) is received from the VAD module 58, parameters relating to jitter, absolute jitter, consecutive positive jitter, packet loss etc are calculated at step 84. These parameters are combined with previously calculated basic parameters in order to calculate level two parameters such as mean, variance, maximum positive value, maximum negative value, sum, difference, running mean, running variance etc. at step 86. For example, level two parameters may include, jitter mean, jitter variance, absolute jitter mean etc.

The level two parameters are combined with previously calculated level two parameters at step 88 in a similar manner to provide level three parameters such as mean, variance, maximum positive value, maximum negative value etc. For example level three parameters may include, maximum positive value of the jitter mean, variance of the jitter variance etc.

Figure 6 illustrates such combination of parameters to provide a final parameter value at step 88. In the example illustrated four basic parameters are combined to provide each level two parameter, three level two parameters are combined to provide a level three parameter.

Finally the level three parameters are combined using a sliding window mechanism which simply sums a predetermined number of previously calculated level three parameters.

This sliding window mechanism is illustrated in Figure 7, where the sliding window sums the previous three level three parameters.

- 5 The calculation of the basic parameter jitter will now be described with reference to Figure 8 which illustrates part of the basic parameterisation of step 84.

Jitter is defined to be the difference between the elapsed
10 time between sending two packets of data and the elapsed time between receiving two packets of data.

Every time new packet data is sent to the parameterisation module 60 a jitter basic parameter is calculated as
15 follows: each packet of data contains a timestamp indicating when the packet was sent. Therefore, elapsed time between sending two packets of data is equal to the packet timestamp minus the previous packet timestamp and is calculated at step 91. Elapsed time between receipt of
20 two packets is calculated using the time of capture recorded by the capture module 50. Therefore elapsed time between receipt of two packets is equal to the packet capture time minus the previous packet capture time and is calculated at step 92, allowing jitter to be calculated
25 from these two values at step 93.

The calculation of the basic parameter short term/long term jitter differential will now be described.

- 30 There are two aspects to jitter, both short term

differences and a long term effects. The jitter buffer 41 in the VOIP gateway 40 (Figure 3) will absorb some of the long term effects of jitter, so these do not necessarily affect the perceived speech quality as much as short term differences. Short-term peaks (or troughs) in jitter have been found to adversely affect speech quality, and therefore a parameter which reflect these short term aspects is very useful for predicting/estimating a MOS.

10 The jitter parameter is used to calculate a long-term average of the jitter. A predetermined adaptation rate P is used, and the long term average (lt_jitter) is calculated at step 94 according to the following equation:

15
$$lt_jitter = (lt_jitter * P) + (abs(jitter) * (1-P))$$

it is worth noting that the absolute value of jitter is used because the size of the difference from a value of zero (no jitter) is important.

20

A differential jitter ($jitter_differential$) parameter is then calculated at step 95 as follows:

$$jitter_differential = abs(jitter) - lt_jitter$$

25

The value of the basic differential jitter (DJ) parameter is then used as described previously to calculate level two parameters such as maximum positive value at step 96, mean value (not shown), variance of the value at step 97; and level three parameters are then calculated such as

30

mean of the maximum positive value at step 98 or mean of the variance of the value at step 99.

5 It will be understood by those skilled in the art that the processes described above may be implemented on a conventional programmable computer, and that a computer program encoding instructions for controlling the programmable computer to perform the above methods may be provided on a computer readable medium.

10

It will be understood that various alterations, modifications, and/or additions may be introduced into the specific embodiment described above without departing from the scope of the present invention.

15